Visualising Personal Data Flows: Insights from a Case Study of Booking.com***

Haiyue Yuan[1][0000−0001−6084−6719], Matthew Boakes[1][0000−0002−9377−6240], Xiao Ma[2], Dongmei Cao[2][0000−0002−2614−3726], and Shujun Li[1][0000−0001−5628−7328]

[1] Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, United Kingdom
{h.yuan-211, m.j.boakes, s.j.li}@kent.ac.uk
[2] Nottingham Business School, Nottingham Trent University, United Kingdom
{xiao.ma, dongmei.cao}@ntu.ac.uk

Abstract. Commercial organisations are holding and processing an ever-increasing amount of personal data. Policies and laws are continually changing to require these companies to be more transparent regarding collection, storage, processing and sharing of this data. This paper reports our work of taking Booking.com as a case study to visualise personal data flows extracted from their privacy policy. By showcasing how the company shares its consumers' personal data, we raise questions and extend discussions on the challenges and limitations of using privacy policies to inform online users about the true scale and the landscape of personal data flows. This case study can inform us about future research on more data flow-oriented privacy policy analysis and on the construction of a more comprehensive ontology on personal data flows in complicated business ecosystems.

Keywords: Personal data, Data flow, Privacy policy, Data sharing, Travel

1 Introduction

Despite the existence of information security policies and data protection laws such as the EU's GDPR (General Data Protection Regulation), over-collection and breach of personal data are constantly happening in the online world. Such data privacy and security issues are partly due to the complex nature of data

* The full edition of this paper can be found on arXiv.org as a preprint at: https://arxiv.org/abs/2304.09603.

** Cite the paper as follows: Haiyue Yuan, Matthew Boakes, Xiao Ma, Dongmei Cao and Shujun Li, "Visualising Personal Data Flows: Insights from a Case Study of Booking.com," in *Intelligent Information Systems: CAiSE Forum 2023, Zaragoza, Spain, June 12–16, 2023, Proceedings*, *Lecture Notes in Business Information Processing* (LNBIP), Volume 477, pp. 52-60, 2023, Springer Nature, https://doi.org/10.1007/978-3-031-34674-3_7, the full edition available as an arXiv.org preprint at https://doi.org/10.48550/arXiv.2304.09603.

collection, processing and sharing processes, where multiple parties are involved and the data owners (more formally called "data subjects") often have no clear view of how their personal data flow between different entities. When a data owner uses their social media (e.g., Facebook) account to access an online service, there will be further personal data flows between the social media company (e.g., Meta) and the service provider. However, the biggest data privacy and security threat caused by the complex situation of personal data sharing among multiple parties [12] is not sufficiently studied in the literature. A recent study [7] investigated such complexity in the tourism domain, and suggested that, while collecting and using personal data can result in more appealing tourism offers and more efficient travel, it can also lead to security risks and privacy concerns, thereby discouraging some travellers from sharing their personal data with service providers. To that end, it will help if travellers are made aware of what personal data will be collected and shared with whom for what purposes. A common approach is to present a privacy policy to users, and some past studies looked into privacy policies in different perspectives such as their impact on users' privacy perception, attitude and behaviour [2, 6], automate privacy policy analysis [1, 5, 3], and readability and visualisation [11, 3, 1]. However, the current approaches are fragmented without comprehensively addressing the full scale of personal data collection and sharing activities and data flows generated by such activities. Having this in mind, our main research question is: *Can we extract and visualise personal data flows between data owners and different data-consuming parties from the analysis of a privacy policy?*

To answer this research question, we used Booking.com as an example to obtain in-depth insights with visual aid to understand how Booking.com and other associated organisations collect and use personal data from customers of Booking.com. We present the following contributions: 1) we propose an approach to systematically analysing and reconstructing personal data flows declared in a privacy policy; 2) we report insights about personal data flows derived from privacy policies via a simple data visualisation approach; 3) we have identified the needs to have a more in-depth investigation of privacy policies from other relevant organisations to get a more comprehensive understanding of personal data flows, and lessons learnt from this case study and future research directions.

2 Related Work

The law in many countries, such as the GDPR in EU member states and the UK, requires service providers to supply privacy policies when collecting personal data is involved, and to present such privacy policies in a concise, transparent, intelligible, and easily accessible form. Researchers have investigated privacy policies in relation to consumers of online services from various perspectives such as its impact on privacy concerns and attitudes, and its visualisation and readability. Bracamonte et al. [2] experimentally evaluated the effects of explanatory information on the perception of the results of an automated privacy policy tool. The results indicate that justification information increases behavioural intention and

perceptions of credibility and utility. Kitkowska et al. [9] studied the influence of privacy policy's visual design by conducting an online experiment and revealed that people feel powerless when acknowledging notices and have no choice but to agree. Ibdah et al. [6] studied users' attitudes and opinions regarding privacy rules. The results suggest that the primary motivation for users to read privacy policies are their concerns about (untrusted) service providers. To improve the readability of privacy policies, Reinhardt et al. [11] proposed developing interactive privacy policies based on the concept of nutrition labels. Harkous et al. [5] proposed an automated framework for privacy policy analysis (Polisis) based on deep learning. Similarly, Andow et al. [1] developed PolicyLint, a framework that can automatically generate ontologies from privacy policies through the use of sentence-level natural language processing. A similar approach was proposed in [8], which led to the creation of a website[3] for visualising Android apps' personal data collection and sharing activities.

Furthermore, there is also research work investigating how to best store and manage personal data. For instance, Verbrugge et al. [13] examined the possibility for a "personal data vault society" and the steps necessary to realise this vision. Fallatah et al. [4] reviewed existing work on personal data stores (PDS), which allow individuals to store, control, and manage their personal data. They argued that one of the technical barriers is the data flow management between different parties. In addition, another way to consolidate the understanding of privacy and personal data collection/sharing is to develop graphical models. More recently, a graphical model proposed by Lu and Li [10] can evaluate personal data flows from "me" (a specific user) and values flowing back to "me" to help inform "me" about privacy-benefit trade-offs.

3 Methodology

In this work, we propose constructing possible flows of personal data through the analysis of an online travel service provider's privacy policy. By visually representing a personal data flow graph derived from the privacy policy, we intend to reveal some potentially overlooked details of personal data sharing activities of consumers of the online service provider.We decided to use the privacy policy of Booking.com[4] as a case study based on the following reasons: 1) Booking.com has the highest revenue globally within the online travel market and is the largest online travel agency by booking volume[5]. 2) Booking.com provides a wide range of features and has a close link with many other subsidiaries of its parent company, Booking Holding Inc., therefore being a good case for understanding how personal data are shared between multiple parties. 3) Booking.com deals with their customers' personal data all the time and with large volume due to the na-

[3] http://android-network-tracing.herokuapp.com/
[4] https://www.booking.com/content/privacy.en-gb.html
[5] https://www.researchandmarkets.com/reports/5330849/global-online-travel-market-2022

ture of its business model. This requires its privacy policy to provide more details on how their consumers' personal data are collected, processed and shared.

To better facilitate the personal data flow mapping and visualisation, we adopted a simplified version of the graphical model proposed in [10] with the main aim of establishing the relationships between the following entities of different types: a) '*Person*' entities stand for natural people in the physical world; b) '*Data*' entities refer to atomic (personal) data items about one or more person entities; c) '*Service*' entities refer to different physical and online services that serve people for a specific purpose; d) '*Organisation*' entities refer to organisations that relate to one or more services. We analysed the privacy policy from the perspective of how data entities flow from users of Booking.com to different data-consuming entities including Booking.com and other organisation entities. More specifically, we analysed the privacy policy from the following two main perspectives: 1) **data collection** is about how Booking.com can implicitly and explicitly collect personal data from its customers, and how Booking.com may receive personal data about its customers from other sources indirectly (i.e., not from its customers directly); and 2) **data sharing** is about how Booking.com shares personal data collected with third parties, including within Booking Holdings Inc. and its other subsidiaries, and with other third parties and online social media service providers. By manually noting down the relationships between different entities while going through the whole privacy policy, we were able to derive a graphical representation of possible personal data flows and a visualisation of the graph.

It is worth noting that the graph presented in this paper is a simplified version, which is based on the assumption that the booker, referring to an individual who arranges a travel booking, is also the sole traveller. It is important to acknowledge that the personal data flows and the data flow graph can be more complicated when the booker is not a traveller or a member of a large group of travellers. For instance, when the booker is an employee of a company or travel agent or a friend/family member of the traveller(s) who does not participate in the booked travel, personal data of both the booker and of the traveller(s) will need to be shared with Booking.com. In such scenarios, personal data flows between the booker and the traveller(s) also need considering, which may also include the case that some traveller(s) may have an account on Booking.com while others do not have one. We consider such more complicated cases out of the scope of this study, and leave them as part of our future work. We also would like to highlight that, although the results presented in this paper are mainly based on the analysis of the privacy policy, we also made dummy bookings on Booking.com to recover and clarify some less apparent information to consolidate our understanding of personal data flows associated with a travel booking made using Booking.com.

4 Results

Figure 1 shows the reconstructed personal data flow graph through the analysis of the privacy policy of Booking.com in terms of personal data shared by consumers of Booking.com. As indicated by the green arrows at top of the graph, the personal data flows from the left to the right side demonstrate how Booking.com can collect their customers' personal data, what types of personal data can be collected, and to what extent Booking.com shares collected personal data with other third parties and for what purposes. In addition, we use 'Challenge 1', 'Challenge 2' and 'Challenge 3' in Figure 1 to represent three main challenges that we have identified, when identifying personal data flows based on analysis of a privacy policy.

4.1 How Booking.com Collects Personal Data

As illustrated in Figure 1, Booking.com can collect their customers' personal data from various sources in several ways. We categorise them into two groups based on the original data controller (the organisation who collects personal data in the first place): *direct data collection* and *indirect data collection*. It is worth noting that all personal data types listed in *Data Boxes A, B, C, D* in Figure 1 are extracted from examples in the privacy policy. However, we acknowledge that it is not an exhaustive list of the personal data types Booking.com may collect. We add three dots at the bottom of *Data Box D* to indicate such incompleteness and also consider this as one of the challenges (i.e., Challenge 1 in Figure 1), which deserves further studies. We provide more details throughout the rest of this section.

1) *Direct data collection* refers to the case that Booking.com collects some personal data directly from their consumers. We further identified two approaches to direct data collection. a) *Explicit direct data collection* means that a person (the traveller or their assistant/helper) provides personal data to Booking.com directly via its website or mobile app as illustrated in Figure 1. Data Box A contains personal data that could be collected specifically for the booking purposes such as the booker's names, telephone number, and email address. Data Box B includes information of the traveller such as their name, date of birth, email address, and dietary requirements. b) *Implicit direct data collection* means that, without explicit data input from a user, Booking.com automatically collects some personal data, which is occurring simultaneously while a user is using Booking.com's website or mobile app. As depicted in *Data Box C* in Figure 1, such personal data could include behavioural data of the user when using a mobile device, the website or the mobile app, the user's social media data, IP addresses, and language settings on the devices, which can be automatically collected using different technologies such as web tracking technologies, web cookies, device sensors, and cross-device tracking technologies. As mentioned in Section 2, substantial work has been done by Jin et al. [8] to extract and visualise mobile apps' data flows[6].

[6] http://android-network-tracing.herokuapp.com/

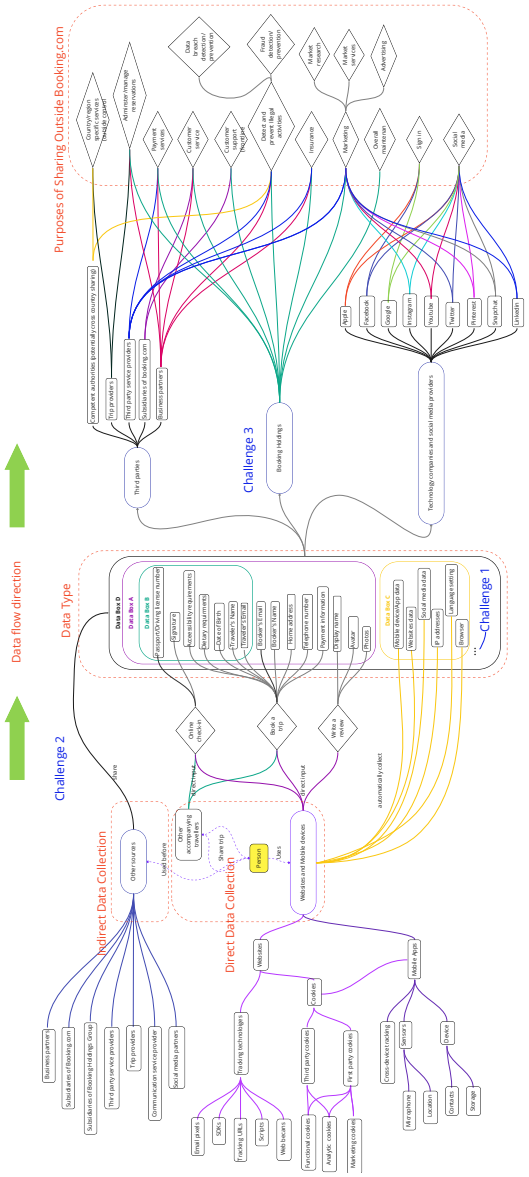

Fig. 1. Booking.com personal data flows diagram extracted from its privacy policy

2) *Indirect data collection* refers to personal data that are not collected by Booking.com directly due to a booker's use of its website or mobile app, but shared with Booking.com by other parties such as third-party service providers, associated business partners, and other subsidiaries of Booking Holdings Inc.

Please note that these parties' privacy policies could have explicitly allowed sharing of their consumers' personal data with Booking.com. Thereby, as shown in *Data Box D* in Figure 1, a user's personal data could be shared with Booking.com and be used together with other personal data directly collected by Booking.com to better serve the person. Here, we would like to highlight that *Data Box D* may contain a much more comprehensive range of personal data, since Booking.com can still gain access to many personal data of their customers indirectly from other third parties. As stated in Booking.com's privacy policy, it is worth noting that how and what other third-party organisations share personal data with Booking.com depend on their business needs and privacy policies. In other words, it is impossible to get more insights about such personal data without analysing privacy policies from other third-party organisations, which is needed to recover the full scale of indirect data collection. We consider this as another challenge (i.e., Challenge 2 in Figure 1) for future research.

4.2 How Booking.com Shares Personal Data

Booking.com claims to collaborate with various organisations to provide more satisfactory services to their users and to serve other legitimate purposes, which often require sharing personal data between such collaborative organisations. By analysing the privacy policy of Booking.com, we have identified three main destinations for data sharing: Booking Holdings Inc. and its subsidiaries, third-party organisations, technology companies and social media providers.

1) Third-party organisations: There are different types of third parties according to Booking.com's privacy policy, and each type has its specific purpose of utilising personal data shared by Booking.com to fulfil users' booking activities. For instance, personal data can be shared with national or local authorities for legal requirements or legal compliance purposes; and Booking.com outsources its administration/reservation services and customer services to a business partner, which would require sharing some personal data to facilitate such outsourced services.

2) Booking Holdings Inc. and its subsidiaries: Being a subsidiary of Booking Holdings Inc., Booking.com can share its users' personal data upward to its parent company. Booking Holdings Inc. provides travel-related services to people in more than 220 countries and territories through many subsidiaries such asPriceline, Agoda, Rentalcars.com, and Momondo[7]. These subsidiaries of Booking Holdings Inc. offer some essential services to each other, and personal data collected by Booking.com may be further spread to other subsidiaries in order to provide more sophisticated combinations of services, as shown in Figure 1, including payment services, admin/manage reservations, customer support, illegal activity detection and prevention.

3) Technology companies and social media providers: Booking.com can share their users' personal data with technology companies and social media service providers in exchange for their services or to provide extra customer benefits.

[7] https://www.bookingholdings.com/about/factsheet/

For example, Booking.com allows customers to sign in using their Apple, Facebook or Google credentials. In addition, using social media features such as integrating social media plugins into Booking.com's website or mobile app and social media messaging services can lead to exchanges of personal data between Booking.com and social media service providers (e.g., Instagram, YouTube, and Twitter). Moreover, another important purpose of such personal data sharing is to conduct market research and to provide more personalised market and advertising services.

However, Booking.com's privacy policy does not contain any information about which type of personal data are shared with which other organisations in detail, preventing us from achieving a full understanding of the landscape of personal data collection and sharing. Thereby, we envisage that another challenge (i.e., Figure 1 Challenge 3) for future work is to conduct more work to study cross-organisational data sharing to consolidate our understanding.

5 Further Discussions and Conclusion

The paper presents a case study to understand how Booking.com collects and shares personal data based on analysis of its privacy policy. By producing a personal data flow graph as a visual aid, we were able to reveal how Booking.com can collect personal data and shares such data with other organisations. Although our work focuses on Booking.com as a case study, the following lessons learnt are likely true for many other online services regarding the challenges of refining privacy policies to reflect the landscape of personal data collection and sharing: 1) the lack of a comprehensive description of the types of personal data that could be collected directly or indirectly; 2) an incomplete description of how and to what extent other organisations can share personal data with online service providers; 3) an unclear description of how and to what extend online services can share their customers' personal data with other third parties; and 4) an unclear disclosure on how personal data collected are used. The lack of clarity and transparency makes it difficult for users to understand the full extent of personal data collection and sharing and how their personal data may be used, therefore subsequently harming their confidence in continuously accepting the business-centric approach to personal data management of online services' users [14].

Furthermore, this study has the following limitations, and we intend to address these in our future work: 1) we only considered using the privacy policy as the main data source to derive the personal data flows; 2) some more fine-grained details of personal data flows are not presented to avoid over-complicating the personal data flow graph; and 3) a user study should be conducted to validate our approach. Last but not the least, we consider our work as the basis to establish a more general approach to automating extraction of personal data flows for any given online services. We hope that this case study can be a stepping stone to elicit more follow-up work on privacy policy analysis, personal data flows, and related graphical modelling and ontological research.

Acknowledgements

The authors were supported by the Engineering and Physical Sciences Research Council, UK Research and Innovation (UKRI), as part the project "PriVELT: PRIvacy-aware personal data management and Value Enhancement for Leisure Travellers", under the grant numbers EP/R033749/1 and EP/R033609/1.

References

1. Andow, B., Mahmud, S.Y., Wang, W., Whitaker, J., Enck, W., Reaves, B., Singh, K., Xie, T.: PolicyLint: Investigating internal privacy policy contradictions on Google Play. In: Proceedings of the 28th USENIX security symposium. pp. 585–602. USENIX Association (2019), https://www.usenix.org/conference/usenixsecurity19/presentation/andow
2. Bracamonte, V., Hidano, S., Tesfay, W.B., Kiyomoto, S.: Effects of explanatory information on privacy policy summarization tool perception. In: Information Systems Security and Privacy: 6th International Conference, ICISSP 2020, Valletta, Malta, February 25–27, 2020, Revised Selected Papers. pp. 156–177. Springer (2022). https://doi.org/10.1007/978-3-030-94900-6_8
3. Carlsson, R., Heino, T., Koivunen, L., Rauti, S., Leppänen, V.: Where does your data go? comparing network traffic and privacy policies of public sector mobile applications. In: Information Systems and Technologies: WorldCIST 2022, Volume 1. pp. 214–225. Springer (2022). https://doi.org/10.1007/978-3-031-04826-5_21
4. Fallatah, K.U., Barhamgi, M., Perera, C.: Personal data stores (PDS): A review. Sensors **23**(3) (2023). https://doi.org/10.3390/s23031477
5. Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K.G., Aberer, K.: Polisis: Automated analysis and presentation of privacy policies using deep learning. In: Proceedings of the 27th USENIX Security Symposium. pp. 531–548. USENIX Association (2018), https://www.usenix.org/conference/usenixsecurity18/presentation/harkous
6. Ibdah, D., Lachtar, N., Raparthi, S.M., Bacha, A.: "why should I read the privacy policy, I just need the service": A study on attitudes and perceptions toward privacy policies. IEEE Access **9**, 166465–166487 (2021). https://doi.org/10.1109/ACCESS.2021.3130086
7. Ioannou, A., Tussyadiah, I., Miller, G.: That's private! understanding travelers' privacy concerns and online data disclosure. Journal of Travel Research **60**(7), 1510–1526 (2021). https://doi.org/10.1177/0047287520951642
8. Jin, H., Liu, M., Dodhia, K., Li, Y., Srivastava, G., Fredrikson, M., Agarwal, Y., Hong, J.I.: Why are they collecting my data? inferring the purposes of network traffic in mobile apps. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(4) (2018). https://doi.org/10.1145/3287051
9. Kitkowska, A., Warner, M., Shulman, Y., Wästlund, E., Martucci, L.A.: Enhancing privacy through the visual design of privacy notices: Exploring the interplay of curiosity, control and affect. In: Proceedings of Sixteenth Symposium on Usable Privacy and Security. USENIX Association (2020), https://www.usenix.org/conference/soups2020/presentation/kitkowska
10. Lu, Y., Li, S.: From data flows to privacy-benefit trade-offs: A user-centric semantic model. Security and Privacy **5**(4) (2022). https://doi.org/10.1002/spy2.225

11. Reinhardt, D., Borchard, J., Hurtienne, J.: Visual interactive privacy policy: The better choice? In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM (2021). https://doi.org/10.1145/3411764.3445465
12. Such, J.M., Criado, N.: Resolving multi-party privacy conflicts in social media. IEEE Transactions on Knowledge and Data Engineering **28**(7), 1851–1863 (2016). https://doi.org/10.1109/TKDE.2016.2539165
13. Verbrugge, S., Vannieuwenborg, F., Van der Wee, M., Colle, D., Taelman, R., Verborgh, R.: Towards a personal data vault society: an interplay between technological and business perspectives. In: Proceedings of the 2021 60th FITCE Communication Days Congress for ICT Professionals: Industrial Data – Cloud, Low Latency and Privacy. IEEE (2021). https://doi.org/10.1109/FITCE53297.2021.9588540
14. Wieringa, J., Kannan, P., Ma, X., Reutterer, T., Risselada, H., Skiera, B.: Data analytics in a privacy-concerned world. Journal of Business Research **122**, 915–925 (2021). https://doi.org/10.1016/j.jbusres.2019.05.005